Scand J Work Environ Health 13 (1987) 486—492

Discovering carcinogens in the occupational environment

Methods of data collection and analysis of a large case-referent monitoring system

by Jack Siemiatycki, PhD,[1, 2] Sholom Wacholder, PhD,[2, 3] Lesley Richardson, MSc,[1] Ron Dewar, MSc,[1, 2] Michel Gérin, PhD[4]

SIEMIATYCKI J, WACHOLDER S, RICHARDSON L, DEWAR R, GÉRIN M. Discovering carcinogens in the occupational environment: Methods of data collection and analysis of a large case-referent monitoring system. *Scand J Work Environ Health* 13 (1987) 486—492. A multi-cancer site, multi-factor, case-referent study was undertaken to generate hypotheses about possible occupational carcinogens. About 20 types of cancer were included. Incident cases among men aged 35—70 years and diagnosed in any of the major Montreal hospitals were eligible. Probing interviews were carried out for 3 726 eligible cases. The interview was designed to obtain detailed lifetime job histories and information on potential confounders. Each job history was reviewed by a team of chemists who translated it into a history of occupational exposures. These occupational exposures were then analyzed as potential risk factors in relation to the sites of cancer included. For each site of cancer analyzed, referents were selected from among the other sites in the study. The analysis was carried out in stages. First a Mantel-Haenszel analysis was undertaken of all cancer-substance associations, stratifying on a limited number of covariates, and, then, for those associations which were noteworthy in the initial analysis, a logistic regression analysis was made taking into account all potential confounders. This report describes the fieldwork and analytical methods.

Key terms: carcinogenesis, epidemiologic methods, neoplasms, occupational diseases.

Of the thousands of substances to which workers are exposed in the workplace, only a handful have ever been assessed in epidemiologic studies for carcinogenic potential (18). It is reasonable to suspect that many occupational carcinogens have not yet been discovered. Thus it is important to institute systematic monitoring procedures to generate data on possible cancer-exposure associations. We previously outlined an approach consisting of a population-based case-referent study covering many sites of cancer (13, 15). The data collection involves obtaining detailed lifetime job histories from subjects and translating each job history into a history of occupational exposures. This last step is accomplished by a team of chemists and hygienists who consider the details and idiosyncracies of each job description and assign exposures in accordance with all available information about each unique situation. The occupational exposures thus inferred become factors for analysis in relation to cancer.

[1] Epidemiology and Preventive Medicine Research Center, Institute Armant-Frappier, Laval-des-Rapides, Québec, Canada.
[2] Department of Epidemiology and Biostatistics, McGill University, Montreal.
[3] Current address: National Cancer Institute, Bethesda, Maryland.
[4] Department of Occupational and Environmental Health, University of Montreal, Montreal, Canada.

Reprint requests to: Dr J Siemiatycki, Centre de recherche en épidémiologie et médecine préventive, Institut Armand-Frappier, 531 boul des Prairies, Laval-des-Rapides, Québec, Canada, H7V 1B7.

Many sites of cancer have been included in the study. For each patient, information has been obtained concerning past exposure to about 300 substances. The overall purpose is to analyze the association between each type of cancer and each substance, as well as to examine associations between cancer types and jobs and industries.

Because of the broad scope of the project and its findings, it has been necessary to analyze, consider, and present subsets of results separately. In fact we have adopted a strategy of subdividing the substances on our checklist into small groups which have some common chemical and/or physical and/or use pattern characteristics and focusing the analyses on each subgroup one at a time. The first was an analysis of nine organic dusts (16), the second an analysis of 12 petroleum-derived liquids (14), and so on. Because of the complexity of the data collection and analysis of this project, we feel it would be useful to present herein the methods used to collect and analyze the data and some general principles of the interpretation of the results. This report will thus be referenced in a series of substantive articles which will follow.

Cases

About 20 sites of cancer were selected for study. The target population consisted of men aged 35—70 years and resident in the area of Montreal (population 2.7 million). The cases were ascertained via 19 major hospitals in the area. These hospitals collectively report

to the Quebec Tumor Registry over 97 % of all tumor diagnoses from the Montreal area. Because of the national health insurance system which provides free medical care, the lack of any major "competing" health care facilities within several hundred kilometers, and the excellent reputation of local medical facilities, we doubt if more than a small fraction of truly eligible cases would be diagnosed outside of the hospitals in our ascertainment system. Diagnoses had to be histologically confirmed for a case to be eligible.

The cases eligible for this study were first diagnosed in September 1979 — June 1985. However, because of limited interviewer resources, we did not ascertain all sites of cancer without interruption during this period. Each summer, there was a two- to four-month period during which incident cases were not entered into the study, so as to allow the interviewers time to complete the previous year's backlog. Furthermore, lung cancer was excluded in the second, third, and sixth years. Rectal cancer was excluded in the first and second year. Prostate cancer was excluded for part of the fourth year and all of the fifth year.

Active regular contact with hospital pathology departments provided rapid case notification. Altogether, 4 576 eligible cases were ascertained. An interviewer visited the patient in the hospital or at home, as required. Approximately half the patients were still in the hospital when first contacted by the interviewer, and the rest had been discharged or were diagnosed as out-patients. Completed interviews or questionnaires were obtained for 3 726 subjects (81.4 %). Reasons for nonresponse were patient refusal (8.0 %); patient death, no next-of-kin found (5.9 %); patient discharged, no valid address available (3.6 %); psychiatric case or unable to speak French, English, or Italian (0.9 %); physician refusal (0.1 %). Table 1 shows the numbers of eligible cases for each site which had over 40 incident eligible cases, and it shows the numbers of patients successfully interviewed. Among the various types of cancer, the response rates varied from 64 to 91 %, most being between 80 and 84 %. To assess the likelihood of nonresponse bias, we examined the similarity between respondents and nonrespondents with respect to several variables which were available from the medical records. The differences were small (11). Eighty-one percent of the completions were obtained from the subjects themselves. The rest were obtained from the next-of-kin of deceased subjects, usually spouses. Face-to-face interview was the strategy of choice, and 82 % of the completions were obtained this way. However, telephone interviews and specially designed self-administered forms were used for hard-to-interview subjects, and these media provided 10 and 8 %, respectively, of the completions.

For each cancer patient, the diagnosis, abstracted from the medical records, was coded by topography and morphology according to the International Classification of Diseases (ICD). The cases could be grouped in many ways. For the most part we used the three-digit level of the ICD topography axis to define the case series. There were some exceptions, however, where numbers warranted subdivision.

We analyzed each of the three topographic subcategories of colorectal cancer as case series, ie, colon (excluding sigmoid), sigmoid colon plus rectosigmoid junction, and rectum. We also analyzed each of four histological subcategories of lung cancer as case series, ie, oat-cell carcinoma, squamous-cell carcinoma, adenocarcinoma, other histological types (including unspecified morphology). The 78 cases which were diagnosed with primary tumors at two different sites during the period of our study have been included in both relevant case series.

We thus collected data on 3 726 cancer patients distributed among many sites of cancer.

Table 1. Sites with over 40 eligible cases of cancer, the numbers ascertained, the numbers of cases for which an interview was carried out, and the response rate.[a]

Cancer sites[b]	Number of months of ascertainment	Number of cases ascertained	Number of cases interviewed	Response rate (%)
Esophagus (150)	49	136	107	78.7
Stomach (151)	49	312	250	80.1
Colon (153, except 153.3)	49	435	364	83.9
Rectosigmoid (153.3, 154.0)[c]	49	285	233	81.8
Rectum (154, except 154.0)	31	215	190	88.4
Liver (155)	49	78	50	64.1
Pancreas (157)	49	164	117	71.3
Lung (162)	26	1 069	857	80.2
Prostate (185)	35	553	452	81.7
Bladder (188)	49	576	486	84.4
Kidney (189)	49	230	181	78.7
Melanoma of the skin (172)	49	147	121	82.3
Hodgkin's lymphoma (201)	49	58	53	91.4
Non-Hodgkin's lymphoma (200, 202)	49	244	206	84.4

[a] In addition to patients with cancers of the sites listed, patients with the following cancers were interviewed: gallbladder, peritoneum, pleura, testes, penis, myeloma. While these could be used in the pool of referents, there were too few of each series to provide reasonable statistical power as case series.
[b] Code of the International Classification of Diseases in parentheses.
[c] Includes sigmoid colon and rectosigmoid junction.

Referents

Our primary strategy was to compare each case series with a reference group drawn from among the other sites. If a substance is not associated with any site or if it is associated with only one site of cancer, then the intercancer strategy should, on the average, give accurate estimates of relative risk. However, if a substance is associated with many sites of cancer, then the excess risk estimates may be too low. Although some carcinogens may act at multiple sites, this possibility has not been found to be a universal phenomenon (9). Even if a chemical produces cancer at multiple sites, it would have to do so to the same degree to slip through the monitoring system. For example, suppose a chemical increases the risk of bladder, liver, and pancreas cancers by factors 5, 3 and 2, respectively, and does not affect other sites. Then an analysis of bladder cancer with the use of a variety of sites of cancers, including liver and pancreas as referents, would yield an observed relative risk lower than the real five but probably high enough to be detected as a health hazard. Intercancer comparisons provide our prime strategy of analysis because they are efficient and because various sources of bias, namely, nonresponse bias and information bias, are minimized. It is a conservative strategy which may attenuate risk estimates, but not exagerate them. We also interviewed a small population reference series and have used that group as an additional reference group. These data have however not yet been fully processed and will not be further considered in the initial series of reports.

For each case series the reference group consisted of all other cancer patients with the following exceptions. Lung cancer cases were excluded from all reference series. The strength of the association between lung cancer and cigarette smoking makes adequate statistical adjustment difficult; errors in measurement or in the functional form of the adjustment procedure could result in residual confounding. In addition we excluded any other site which was anatomically contiguous (eg, esophagus for stomach and vice versa) from the reference pool for a given site. This exclusion was made because there might be misclassification around the junction and because, in some instances, there is a greater likelihood of shared etiologies. In addition, for the three sites which were not ascertained in certain years of the study, lung, rectum and prostate, the referents consisted of other subjects who entered the study in the same years as the corresponding cases. This practice was initiated to control for any variation in the quality of the interviews or exposure assessment across the years of the study, despite our efforts to insure consistency. Finally, subjects with two primary cancers who were included in two case series were not allowed to serve as referents for themselves.

Table 2 shows, for each of the sites analyzed as case series, the sites that served as reference and the numbers of subjects included in the analyses.

Table 2. Types of cancer analyzed as case series, sites excluded from the reference series, and numbers of cases and referents.

Site of the cancer case series	Cancer sites excluded[a] from the reference series	Number of cases	Number of referents
Esophagus	Lung, stomach	107	2514
Stomach	Lung, esophagus	250	2514
Colon	Lung, other colorectal	364	2081
Rectosigmoid	Lung, other colorectal	233	2081
Rectum	Lung, other colorectal	190	1315
Liver	Lung	50	2806
Pancreas	Lung	117	2741
Lung			
Oat cell	Other lung	159	1523
Squamous cell	Other lung	359	1523
Adenocarcinoma	Other lung	162	1523
Other cell types[b]	Other lung	177	1523
Prostate	Lung	452	1733
Bladder	Lung, kidney	486	2196
Kidney	Lung, bladder	181	2196
Melanoma of the skin	Lung	121	2737
Hodgkin's lymphoma	Lung, other lymphoma	53	2599
Non-Hodgkin's lymphoma	Lung, Hodgkin's	206	2599

[a] For each case series, all cancer patients interviewed served as referents with the exceptions listed in this column. Furthermore, for rectum, lung and prostate, only those subjects interviewed during the same ascertainment periods as the three respective site series were used as referents.

[b] This is a heterogeneous grouping which includes large cell, spindle cell, adenosquamous and "carcinoma, not otherwise specified."

Data collection

The questionnaire had the following two parts: (i) a structured section requesting information on important potential confounders (eg, age and ethnic group; residential history; schooling; home environment such as public water supply, home heating and cooking facilities; current and childhood socioeconomic status, consumption of cigarettes, alcohol, coffee, tea; hobbies; consumption of foods containing carotene; height and weight) and (ii) a semi-structured probing section designed to obtain a detailed description of each job the subject had had in his working lifetime. The interviewers were trained to probe for as much information as the patients could supply on the company's activities, the raw materials used by the company, the final product, the machines used, the worker's responsibilities for machine maintenance, the type of room or building in which the work took place, activities of workmates, presence of gases, fumes or dusts, and any other information which could furnish a clue as to possible chemical or physical exposures incurred by the subject. The validity of reported job histories was assessed and found to be adequate (2).

A team of chemists and hygienists working with us had the responsibility of examining each completed questionnaire and translating each job into a list of potential exposures. They used a checklist which explicitly listed some 300 of the most common occupational exposures in Montreal. The checklist, as well as the coding procedures, have been described by Gérin et al (5). For each product thought to be present in each job environment, the chemists noted their confidence that the exposure actually occurred (possible, probable, definite), the frequency of exposure during

a normal workweek (<5, 5—30, >30 %), and the concentration of the agent in the environment (low, medium, high). The designation of low, medium, or high for concentration was determined with reference to certain occupations in which the substance occurs. The score has no meaning on any absolute scale, nor are the levels comparable between substances. The dates of the beginning and ending of each job were recorded, and thus of the corresponding exposures in each job. The jobs themselves and the industries were coded according to standard Canadian classifications.

The team of chemists relied on the following sources as a basis for their retrospective exposure assessment: their own industrial experience and chemical knowledge, old and new technical and bibliographic material describing industrial processes, consultations with experts familiar with particular industries, and, of course, previously coded files of the same job category. The interrater agreement was assessed and found to be adequate (6). The chemists were unaware of the site of the patient's cancer.

Defining exposure groups

For each substance the estimation of odds ratios required the definition of an unexposed group and of one or more levels of exposure. In the basic analyses, "unexposed" was a substance-specific notion. For instance, in the analysis of gasoline, the "unexposed" were unexposed to gasoline, but they may have been exposed to other substances. Since the only intended difference between "exposed" and "unexposed" was in exposure to gasoline, the results provided specific evidence concerning risks due to gasoline. If by chance there were differences between the two groups in the distribution of exposure to other substances, these could, in theory, be adjusted for in the analysis.

In the analyses outlined below, "exposed" was sometimes defined as any exposure and sometimes according to a more stringent criterion based on degree and duration of exposure. For each substance to which a subject was thought to have been exposed, we had the following four dimensions of information: (i) concentration of exposure, (ii) frequency of exposure, (iii) confidence that the exposure occurred, and (iv) duration. Duration was available in years, while the three others dimensions were coded on three-point ordinal scales. To simplify the analyses, it was useful to create a synthetic index of cumulative exposure based on concentration, frequency, confidence, and duration. For some analyses, this index of cumulative exposure was divided by duration to provide an index of the average level of exposure. While the categories of each dimension were simply coded 0, 1, 2 or 3, these ordinal values did not represent the relative weightings of concentration, frequency, and confidence as the chemists actually used them. It was impossible to be precise, but we felt that the square of these terms more accurately reflected the way they were used. This can be seen more clearly in the case of the frequency scale, on which the absolute values corresponding to the three categories (<5, 5—30, >30 %) are much steeper than 1: 2: 3. Although not as easily quantified as the frequence scale, the difference between concentration levels of 3 and 1 would also reflect differences in actual values closer to ninefold than to threefold. Thus, before the exposure indices were created, the four-point components were recoded as 0, 1, 4, and 9.

For each substance, i, and each year of the subject's career, j, the average level of exposure was defined as

$$X_{ij} = \text{concentration}_{ij} \times \text{frequency}_{ij} \times \text{confidence}_{ij}.$$

The cumulative exposure to substance i was defined as

$$E_i = \sum_j X_{ij},$$

j being summed over the years of exposure. The average level of exposure to substance i was defined as:

$$A_i = E_i/\text{duration}_i,$$

where duration signified the number of years of exposure to this substance.

Mantel-Haenszel screening analyses

There was information on scores of nonoccupational factors and hundreds of occupational exposures. In the analysis of the association between a given substance and a given type of cancer, potential confounders included the nonoccupational factors and the occupational exposures apart from the one under study. To reduce the problem to manageable dimensions, we carried out several stages of analysis.

For each set of substances forming the subject of a report, the first stage was an attempt to screen all cancer × substance associations to pick up those which warranted further attention. It is not self-evident whether it is optimal to base this screening analysis on a dichotomization of any exposure versus no exposure or of substantial exposure versus no exposure. It depends on the numbers exposed at each level and the shape of the dose-response curve, if there is an association. To improve the chance of detecting true associations, we carried out the initial screening analyses once with the exposure dichotomy defined as any exposure versus none and then as substantial exposure versus none. Substantial exposure was defined on the basis of the cumulative exposure index E_i and was based on a cut point determined by the median of all nonzero E_i scores.

We selected certain potential confounders as a priori confounders to be included in every analysis. These were ethnic group, age, socioeconomic status as measured by self-reported income, cigarette smoking, and the dirtiness of the jobs held. This last variable represents an attempt to distinguish "clean" work histories from "dirty" ones, or white-collar from blue-collar ones. This procedure was based on an evaluation, by

our team of chemists, of the dirtiness of the job corresponding to each four-digit job category in the Canadian occupational classification system. Each job was scored from 0 to 6. Thus, for each year of the subject's career, we derived a score associated with the job title held. The overall dirtiness score was an average of these annual scores over the subject's working lifetime.

The two sets of screening analyses were based on the procedure of Mantel & Haenszel (8). A priori covariates were stratified as follows: ethnicity (French, other), age (35—54 and 55—70 years), socioeconomic class (two categories defined by the median income), dirtiness (two categories defined by the median dirtiness index), cigarette consumption (nonsmoker, 0—600 pack-years, >600 pack-years). A flexible program was written for the purpose of computing large numbers of Mantel-Haenszel estimates in a single run (4). The confidence intervals were computed with the variance estimator suggested by Robins et al (12).

Repeat analysis among French Canadians only

The Montreal area is made up of many ethnic groups. French Canadians account for almost two-thirds of the population, and most of the rest are derived from other European nations. The melting pot phenomenon has been much less imposing in Montreal than in other parts of North America, and the various groups have remained relatively distinctive genetic and cultural entities. In addition different ethnic groups in Montreal have tended to be concentrated in different occupational/industrial spheres. In the past, French Canadians were underrepresented in white-collar jobs, while certain ethnic groups (British, Jewish) were underrepresented in blue-collar jobs. The opportunities for confounding and effect modification are obvious, since the different genetic profiles and cultural habits of the diverse ethnic groups may well have their own effect, and they may interact with occupational exposure in determining cancer risk as well. Furthermore, any strategy for the statistical control of confounding may fail if unverifiable model assumptions are incorrect.

Because of the problems and uncertainties in this regard, we carried out the Mantel-Haenszel analyses twice, once among all subjects, as already described, and once among French Canadians only. The analyses for French Canadians had the advantage of eliminating one important source of confounding, and, because they comprise 60 % of our study subjects, the loss of statistical precision was not too great. The analyses for all the subjects had the advantage of greater numbers and thus greater statistical precision.

Empirical search for confounders

On the basis of the Mantel-Haenszel screening results, associations with elevated odds ratios were selected for further in-depth investigation. Each association thus earmarked underwent an analysis to determine which of the hundreds of available variables (both occupational and nonoccupational) might be confounders and then an analysis to estimate the odds ratio, the confounders being taken into account. The search for confounders was based on the empirical principle of finding those covariates which, when included as stratification variables, changed the estimate of the disease-exposure odds ratio. This approach to selecting confounders is admittedly controversial (3, 7, 10). Starting from the Mantel-Haenszel test which had been earmarked in the first stage of screening, we systematically added one different stratification covariate at a time to the five a priori variables already in the model and compared the new odds ratio estimate with the one based on the a priori confounders only. If the two odds ratios differed by more than 10 % (in either direction), the "extra" covariate was earmarked as a potential confounder. Each association thereby gave rise to a variable number of such potential confounders. For lung cancer we also included some confounders on the basis of prior knowledge, ie, asbestos, chromium, and nickel.

Logistic regression analyses

We investigated three different facets of each selected association. We first estimated the risk associated with any level or duration of exposure to the substance (ie, any versus none). To assess dose-response, we then estimated the risks associated with various exposure levels and/or durations.

The third facet of each association concerned the risks in subgroups which received their exposure to the substance in different occupations. For instance, in assessing the gasoline-stomach cancer association, we estimated the stomach cancer risk among mechanics and repairmen exposed to gasoline, among service station workers exposed to gasoline, and among the other occupationally defined subgroups of men exposed to gasoline. It is useful to know whether the apparent excess risk for a substance, gasoline in this example, is spread across the occupations in which it is found or whether it is concentrated in a single occupation. If we find an overall excess risk for a substance and find that the risk is not concentrated in any particular occupation, then either the substance is a causal agent, or the true causal agent is also found in the same range of occupations but was not included as a confounder. However, if the risk is concentrated in one occupation, then the following interpretations are possible:

1. The substance is the causal agent and occurs in higher dosages in this occupation than in others. If our chemists properly designated the relative level and frequency of exposure in different occupations, then there would be a trend in our dose-response analysis.

If they had erred in the relative levels of exposure given to different occupations, then there may not even have been an apparent dose-response effect.

2. The substance is the causal agent and occurs in a qualitatively and crucially different form in this occupation than in others. The difference may be in the chemical composition, or it may be in the physical circumstances of exposure (eg, temperature, mode of contact, concomittant exposures).

3. The substance may not be a causal agent at all but rather a marker for the true causal agent. This possibility is to some extent detectable in comparisons of the risk estimate for the subset of the occupational group which was not exposed to the substance with that for the subset which was exposed to the substance. If the risks are equal in the two subsets, it is likely that the risk is due to some characteristic of the occupational group other than the substance under scrutiny.

For each set of odds ratios and corresponding standard errors to be estimated (ie, for each facet), a logistic regression model was fit with the generalized linear interactive modeling (GLIM) program (1). Case/referent status was the outcome variable, and the exposure variable was defined differently for each of the three facets. Each model included, in addition to the exposure variable, all other potential variables identified in the empirical search for potential confounders, except for any other substances too closely related to the exposure under study. For instance, the inclusion of the substance "lead compounds" as a confounder for leaded gasoline would constitute overmatching.

The analyses described herein represent the basic framework through which all associations in our data set will pass. Those found to be of interest will undergo further analysis.

Interpretation

Our statistical analysis strategy has the effect of adjusting each odds ratio estimate for hundreds of occupational and nonoccupational factors. Although the final logistic regression model may include only a handful of factors, the preceding steps of our analysis effectively rule out the risk of serious confounding due to any other of the variables at our disposal. Of course, there are practical limitations to this ideal. The validity of such retrospective exposure coding cannot be perfect; errors in exposure coding could lead to false negative results if the substance under scrutiny were seriously miscoded or it could lead to false positive results if an important confounding factor were seriously miscoded.

In any study, whether one is evaluating one association or many, there is a probability of false positives, that is, of falsely rejecting the null hypothesis of no association. For any given association, this probability is the same whether one has investigated only that association or many others as well. However, given that we estimate many odds ratios and in fact estimated each odds ratio under four sets of conditions, it is certain that some will be significant by chance. To help tease out the true from false positives, we subject each "significant" association to in-depth analyses aimed at (i) eliminating the effects of all confounding factors, (ii) establishing whether there is any "dose-response" relationship, and (iii) establishing whether the risk seems to be attributable to the substance or, rather, to an occupation.

We disagree with the claim that evidence from a "hypothesis-generating" study is weaker than evidence from a "hypothesis-testing" study. For any given association the same result would be observed if the study had been set up to investigate only that association or others as well (17). Usually studies to test hypotheses are carried out when there has been previous evidence indicating an association. A positive finding in the new research would then be added to the previous evidence, and the weight of evidence could become convincing. For many associations in a study such as ours, excess risk estimates may represent the first evidence of a putative association and for that reason be less convincing. That is, it is the quality and quantity of evidence on a given association that makes it convincing, not whether the investigator had one or another hypothesis in mind, nor whether he examined many associations simultaneously.

Acknowledgments

This research was supported by grants from the Institut de recherche en santé et sécurité du travail du Québec, the National Health Research and Development Program, and the National Cancer Institute of Canada.

Mr J Pellerin and Mr L Wald established and maintained the data management system. The system for the chemical coding of jobs was developed in collaboration with Mr H Kemper, Mr D Bégin, Mr R Lakhani, Dr L Nadon, Ms P Stewart, and Mr L Laroche. Dr B Terracini provided helpful advice on this manuscript. The manuscript was prepared by Ms L Carfagnini and Ms C Dupuis.

References

1. Baker RJ, Nelder JA. The GLIM System: Release 3. Generalised linear interactive modelling. Royal Statistical Society, Oxford 1978.
2. Baumgarten M, Siemiatycki J, Gibbs GW. Validity of work histories obtained by interview for epidemiologic purposes. Am J Epidemiol 118 (1983) 583—591.
3. Breslow NE, Day NE. The analysis of case-control studies. International Agency for Research on Cancer, Lyon 1980.
4. Dewar RAD, Siemiatycki J. A program for point and interval calculation of odds ratios and attributable risks from unmatched case-control data. Int J Bio Med

Comput 16 (1985) 183—190.
5. Gérin M, Siemiatycki J, Kemper H, Bégin D. Obtaining occupational exposure histories in epidemiologic case-control studies. J Occup Med 27 (1985) 420—426.
6. Goldberg MS, Siemiatycki J, Gérin M. Inter-rater agreement in assessing occupational exposure in a case-control study. Br J Ind Med 43 (1986) 667—676.
7. Grayson DA. Confounding confounding. Am J Epidemiol 126 (1987) 546—553.
8. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22 (1959) 719—748.
9. Merletti F, Heseltine E, Saracci R, Simonato L., Vainio H, Wilbourn J. Target organs for carcinogenicity of chemicals and industrial exposures in humans: A review of results in the IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. Cancer Res 44 (1984) 2244—2250.
10. Miettinen OS, Cook EF. Confounding: Essence and detection. Am J Epidemiol 114 (1981) 593—603.
11. Richardson L, Siemiatycki J. Case-control study methods — When to interview subjects and nonresponse bias. Presented at the Fourth International Symposium on Epidemiology in Occupational Health, Como, Italy, 1985.
12. Robins J, Greenland S, Breslow NE. A general estimator for the variance of the Mantel-Haenszel odds ratio. Am J Epidemiol 124 (1986) 719—723.
13. Siemiatycki J, Day N, Fabry J, Cooper J. Discovering carcinogens in the occupational environment: A novel epidemiologic approach. J Natl Cancer Inst 66 (1981) 217—225.
14. Siemiatycki J, Dewar R, Nadon L, Gérin M, Richardson L, Wacholder S. Associations between several sites of cancer and twelve petroleum-derived liquids: Results from a case-referent study in Montreal. Scand J Work Environ Health 13 (1987) 493—504.
15. Siemiatycki J, Gérin M, Richardson L, Hubert J, Kemper H. Preliminary report of an exposure-based, case-control monitoring system for discovering occupational carcinogens. Teratogenesis Carcinog Mutagenesis 2 (1982) 169—177.
16. Siemiatycki J, Richardson L, Gérin M, Goldberg M, Dewar R, Désy M, Campbell S, Wacholder S. Associations between several sites of cancer and nine organic dusts: Results from an hypothesis-generating case-control study in Montreal, 1979—1983. Am J Epidemiol 123 (1986) 235—249.
17. Thomas DC, Siemiatycki J, Dewar R, Robins J, Goldberg M, Armstrong BG. The problem of multiple inference in studies designed to generate hypotheses. Am J Epidemiol 122 (1985) 1080—1095.
18. US National Research Council. Toxicity testing: Strategies to determine needs and priorities. National Academy Press, Washington, DC 1984.

Received for publication: 12 October 1987